Exhibit

Professional Program Session Record **10**

System Design Flexibility Offered by Flash EPROM/E²PROM

Electronic Show and Convention
November 15-17/Anaheim, California

Sponsored by Los Angeles Council and San Francisco Bay Area Council, IEEE
Southern and Northern California Chapters, ERA

FLASH MEMORIES CHANGE SYSTEM DESIGN FUNDAMENTALS

Saul Zales
Applications Engineer
Intel Corporation
1900 Prairie City Road
Folsom, CA 95630

ABSTRACT

Flash memory will be the main direct-access memory of the future. It possesses all traits required to achieve this stature: high density, high performance (random access, low power, erase/write), proven reliability, low cost and long-term scaling potential. This paper provides an overall perspective of embedded and reprogrammable memory architectures and flash memory's role in changing system design fundamentals.

The first section discusses today's system environment. The second part focuses on alternative semiconductor technologies as they relate to current issues. And the final section shows the next evolutionary step in memory architectures, flash memory, including both embedded control and reprogrammable system applications.

TECHNOLOGY AND MEMORY ARCHITECTURES

Memory architectures evolve based on current fundamentally sound technology. For example, computer systems of the 1960s and early 1970s used ferrite-core memory as their main storage medium. Designers added external memory using magnetic tape, disks, drum, and paper cards. In contrast, today's designs combine DRAM with high-capacity, fixed magnetic disks. The performance and economic advantages of those main and external memory technologies caused basic architectural shifts. Figure 1 shows the memory usage evolution vs. available technology.

TODAY'S ENVIRONMENT -- THE 40,000 FOOT VIEW

Advances in VLSI circuitry have brought mainframe functionality to the office desktop computer. Such a high degree of integration permits system designers to develop products at a fraction of the cost of larger minicomputer or mainframe systems, without sacrificing performance or functionality. Unfortunately, software vendors have repeatedly fallen behind the advances in hardware sophistication.

To achieve similar levels of sophistication, software programs continually grow larger. This trend boosts both the main and auxiliary system memory requirements. For example, in 1981, sixty-four kilobytes of main memory was considered sufficient; today that amount barely holds the operating system. Typically, AT-class personal computers contain one megabyte of DRAM, and 386[TM]-based systems contain two to eight megabytes.

A look at the auxiliary storage market shows a corresponding increase in sophistication. The migration from paper tape to 8-inch floppy drives, to 5.25-inch, to then 3.5-inch drives has occurred at an extremely fast pace. In the fixed disk arena, density has increased from 10MB subsystems offered on the early PC-XTs to the 120MB and higher densities available today. The emerging magneto-optical disks offer the next density generation, at roughly an order of magnitude larger capacity than comparable magnetic disks.

	1st Gen.	2nd Gen.	3rd Gen.	Late 3rd
Electrical Components	Tubes	Transistors	IC	IC
Main Memory	Tubes, Delay lines	Magnetic Drum & Core	Magnetic Core & Mag. Media	Semiconductor RAM
Auxiliary Memory	Paper Tape, Cards & Delay Lines	Magnetic Drum, Paper Cards	2nd Gen., Extended Core & Mag. Media	Same as 3rd Gen.

Figure 1. Computer Generations vs. Memory Architectures

Figure 2 shows the software growth from two perspectives: embedded controller memory limits and reprogrammable operating systems memory requirements.

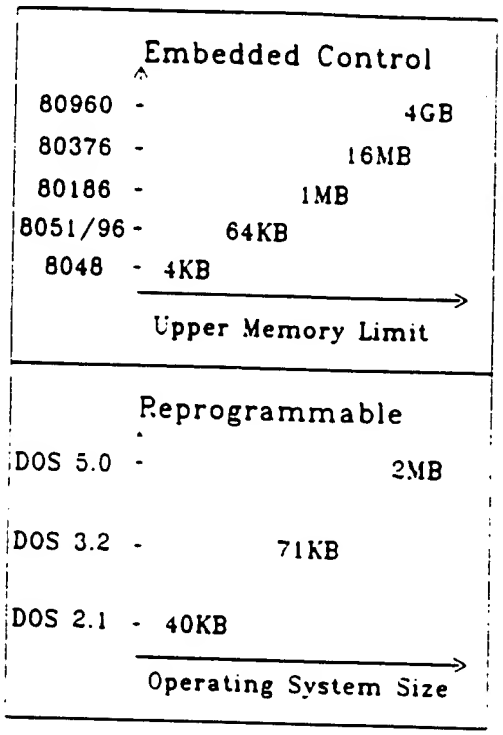

Figure 2.
Software Growth vs. System Complexity

Things Look Great - What's the Impetus for Change?

Despite increased capacity and capability of modern memory systems, they still have performance and reliability drawbacks. In the reprogrammable market, the DRAM plus disk architectural disadvantage is system response time. The flaw in the embedded control environment is DRAM's volatility. Megabytes of memory must be reloaded from auxiliary storage after each power disruption and loaded again each time the application changes. In either case, the disadvantage contributes to a hidden cost of the DRAM/disk architecture.

The File Transfer Bottleneck

There is a well-accepted premise that system response time approaching one second or more significantly affects user performance. Personal computer users encounter many such delays during normal usage. For example, loading a 300KB graphics application program may take 3 - 5 seconds. During program execution, the experienced user sets safety backups to occur regularly. Each backup disrupts the user's keyboard input and potentially his or her train of thought.

Now consider the costs associated with three million dollar time-share systems (mainframes), where CPU interconnect time costs many times more than personal computers. Satisfying all applications requires page-swapping applications to hard drives. During peak usage, page swapping drags the system's efficiency down to 20 percent due to the file transfer bottleneck. That means five other systems would be required to regain the equivalent capability of one system.

Power Failures

Different manifestations of the disk/DRAM problem appear in the embedded control environment. A networked factory hums along nicely most of the year. One day, the power grid in the region fails and with it entire factories. On power-up, every DRAM-based system must re-boot and reload its program code. Few networks handle this strain gracefully, and the factory may take 2-4 hours to get back on-line. During this time, operators sit idle, systems stop producing, and revenue is lost due to missed shipments. Also, an off-line private telephone switching system (PABX), which stores the switch information in DRAM, affects revenue directly -- orders cannot be placed until the DRAM is reloaded!

Additionally, power slumps, also known as brown-outs, wreak havoc on DRAM-based systems. These occur typically during the summer months when air conditioners cause surges in power usage. The problem with DRAM in these situations stems from the technology's dynamic nature. Brown-outs cause systems to run slower and miss DRAM refresh cycles. The DRAM-based system may not cease operation immediately, but probably contains spurious code or data which could cause failure later.

Many designers have chosen battery-backed SRAM as an alternative to DRAM. This commonly-used practice has a number of drawbacks. For starters, the battery will eventually fail.

Additionally, battery-backed SRAM-based systems require extra board space and circuitry for battery charging, power sensing, and supply switching on power fail. This circuitry adds to the total system cost. Once designed into a system, the battery eliminates the possibility of infant mortality screening (burn-in). It also directly limits the temperature operating range of a system. Figure 3 shows characteristics which limit the usefulness of batteries.

Longer on shelf -->	Decreased Amp-hr. capacity, accelerated by temperature
Lower temp -->	Lower output voltage, and faster depletion

Figure 3.

The complex combination of shelf-life, temperature, and current consumption determines a battery's discharge rate and voltage regulation. Power failures may not be predictable.

Traditional EEPROM technology offers designers a nonvolatile alternative. However, the density limitations and relatively high cost of EEPROMs have deterred its acceptance as a mainstream technology.

ALTERNATIVE TECHNOLOGIES AND ARCHITECTURAL CONSIDERATIONS

Density

Beyond the issues of DRAM's volatility associated with embedded and reprogrammable systems, one must look at the main driving force in random access memory architectures -- density and cost. DRAM has become the volume density driver of all solid-state memories, despite its volatile nature. Its read/write capabilities and one transistor/ one capacitor memory-cell architecture drove DRAM to the forefront in terms of both density and cost per bit.

Unlike DRAM, EPROM requires just one transistor per cell. Unfortunately, EPROM is not in-system alterable; hence it lags DRAM in volume and therefore cost per bit. EEPROM on the other hand enables nonvolatile in-system writing, but requires an extra transistor per cell. This cell design coupled with the complexity of the EEPROM process, causes it to lag further and further behind DRAM and EPROM in both cost and density. Similarly, SRAM, with its 4 or 6 transistor configurations lags DRAM and EPROM in both density and cost per bit.

Reliability and Future Densities

According to industry watchers, DRAM currently reigns over other random access memories in usage. Twenty years ago, the forecast for dynamic memory clearly was not as bright. DRAM required uncompromising system-level timing and control to ensure proper operation. Early adapters struggled to meet DRAM's refresh requirements and memory subsystem failures were common.

However, DRAM's simple fabrication process drove its cost down and usage up. Engineers worked around the technology's quirks, and embraced it, first in kilobytes, and today in megabyte chunks. From this questionable beginning arose DRAM controllers, DRAM subsystems, and use of an extra parity bit for every eight data bits. However, shrinking lithographies have attacked DRAM at its single strength -- simple fabrication processes which enabled densities to double every year and a half.

At today's sub-micron lithography, DRAM's charge-storage capacitor no longer holds sufficient charge. Comparatively large alpha particles, emitted from the sun and other sources, bombard the stored electrons and cause soft errors. The memory cells can easily lose enough charge to change states causing erroneous operation. DRAM manufacturers have had to adopt new and more costly fabrication approaches. These include both trench capacitors and multi-poly layered stacked-cell capacitors. Physics authorities question DRAM's future beyond the 0.5 micron lithography and DRAM manufacturers are spending considerable sums to overcome the problems.

Other random-access technologies share similar problems. Four-transistor SRAM offers designers density, but they are increasingly susceptible to alpha particle attack. Six transistor designs offer better alpha particle immunity but compromise on density and therefore raw cost. Another solution, die-coats on four-transistor cells, is also costly.

An architectural feature of traditional EEPROM technologies limits their reliability -- real-time update capability. Real-time updates require fast erasure and programming

mechanisms. To obtain the real-time speeds, an internal voltage of 18V to 30V is used. This voltage produces electric fields on the order of 2,000,000 V/cm greater than those produced at 12V. The tremendous electric field exacerbates small defects in the 100 angstrom oxide (through which tunnelling occurs).

This phenomenon leads to an average 0.3-0.5 failure rate per 1,000 erase/write cycles or 3-5% per 10,000 cycles. (The failure rate is the percentage of parts that exhibit single bit charge loss due to oxide defects caused by cycling.)

UV EPROMs offer the highest reliability of all solid state memories. Also, their simple, oxide-encapsulated storage node provides for many years of future densities. However, as stated earlier, EPROM market growth has not achieved the magnitude of DRAMs due to EPROM's inability to erase in-system.

FLASH MEMORY - THE 40,000 FOOT VIEW

Combining the positive traits of all solid state memories, one can profile the perfect memory: reliable, dense, simple cell structure, nonvolatile yet changeable in-system, and random-access. Intel's ETOX[TM] (EPROM Tunnel Oxide) flash memory technology fulfills this list of requirements. The ETOX process offers the next stage in the evolutionary chain of memory architectures by crossing the volatile and nonvolatile memory boundaries. It also bridges the auxiliary and primary storage media into one.

Further, based on five years of CHMOS-IIE* EPROM manufacturing experience, Intel's flash memory offers excellent reliability. Application code or archival data is assured more than 10 years data retention (100 years typical). Designers need not fear alpha particle exposure, battery power loss, battery chemical leakage, power glitches or any other scourge of modern-day systems. Used in bulk storage designs, flash memory eliminates mechanical wear-out failures.

Another aspect of Intel's flash as a random-access memory devices is its rewrite endurance. Intel's ETOX flash memory has a failure rate less than 0.01% per 100 write cycles and less than 0.1% per 10,000 cycles.

*CHMOS is a patented process of Intel Corporation.

In fact, the failure rate to date on units cycled through 100K cycles is also less than 0.1%.

Like its EPROM predecessor, the ETOX flash memory cell requires one simple transistor per cell. This provides for a rapid advance along the density treadmill. Over time, Intel's flash memory should reach parity with EPROM densities.

RELIABILITY	10 - 100 YRS DATA RETENTION
CELL COMPLEXITY	1 TRANSISTOR
SCALING POTENTIAL	HIGH
NONVOLATILE	YES
IN-SYSTEM UPDATABLE	YES

Figure 4. ETOX[TM] Flash Memory

Additionally, Intel's ETOX technology approach allows for scaling to the 0.1-micron lithography. The floating gate can scale indefinitely due to 1) its inherent quantum-mechanical barrier that maintains the charge level; and 2) the charge on the gate controls the cells threshold. Both of these attributes allow the cell size to scale smaller along with the x and y dimensions. DRAMs, on the other hand, must maintain an absolute amount of charge. As the x and y transistor dimensions shrink, its storage capacitor grows larger and larger. Figure 4 summarizes this phenomenon.

DESIGN FUNDAMENTALS CHANGING

New technologies displace current standards by crossing supposed fundamental boundaries. For instance, the first mechanical engines overcame the limitations of horsepower. The first diodes and transistors crossed the electrical plane into the sublime world of semiconductors. Likewise, the first integrated circuit inventors leapt across the interconnect barrier. Today, system designers face a four-way intersection consisting of primary and auxiliary storage and volatile and nonvolatile semiconductor memory.

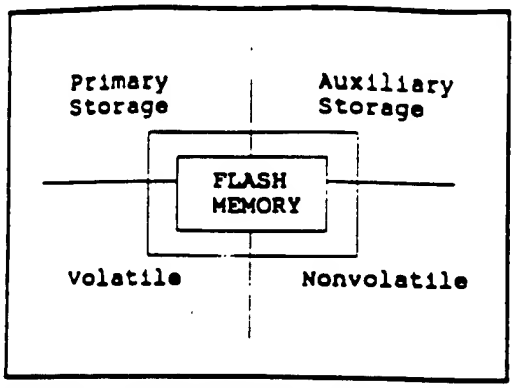

Figure 5. Convergence of Memory Systems

Flash Memory in Embedded Control

Many embedded control applications, such as consumer electronics, do not require reprogramming. The microcontroller's task is simple and straightforward and the code is stored in EPROM. On the other hand, high-end embedded control systems service complex environments. The code for these systems may be revised numerous times in the product's early production stages or may even demand regular updates to fulfill the product's intended purpose. Each update can be accomplished without a service call by using flash memory instead of EPROM.

Also, reconsider the automated factory. Some harsh environments preclude mechanical storage media in each system, so file servers are tucked in a safe corner. Set-up technicians convert from job to job by downloading new code over a network to each system. Alternatively, the manufacturer pays extra for disk drives and controllers in each system. This method works fine until a power failure or brown-out. At that point every piece of equipment must be reloaded. The overwnelming cost benefit of flash over the DRAM-plus-disk combination is the opportunity of factory re-setup time after power loss.

Another example shows this point in a stronger light. Reconsider the PABX -- the lifeblood of some businesses. Many PABX systems still use paper or magnetic tape to store the switch settings. After a power outage, an operator must re-boot and serially reload the switch's DRAM serially from the tape. In this case, the benefit of flash memory is the zero-downtime once power is restored: Sales clerks can return to business as normal with no further delay.

Flash Memory in the Reprogrammable World

Similar to most secondary storage systems, flash memory is block-erasable. This fact precludes its use from data manipulation tasks which require byte-alterability. But this is not a significant drawback, considering the habits of the average computer user. Advanced systems today allow for megabytes of volatile memory. The user loads one to four applications per day from a mechanical auxiliary storage sub-system. This sub-system consumes a couple watts of power, adding to the system weight, and decreasing its ruggedness. While this may be acceptable for bulky desktop computers, it is unacceptable for the exploding portable market. On each application change, the user must wait for the small disk drive to locate the new application and transfer it serially from the disk to the system bus.

Imagine a system where the same memory can serve as both application storage and be random-access. The user calls for a word processing program and it appears instantly. The user then decides to retrieve data from a spreadsheet program. It too is available instantly.

A solid-state system with this performance can be built today using battery-backed SRAM or EEPROM. However, the designer using 2, 4 or 6 transistor technologies incurs the density and cost penalties. With SRAM, the user also faces the nuisance of maintaining battery levels to assure data retention. A flash memory-based system avoids these drawbacks. With flash, the only system design consideration is a regulated 12V power supply to store. Designers can use inexpensive 5V to 12V local power supplies for this task.

Flash memory in the reprogrammable environment is a revolutionary change for system architectures. It will require software as well as hardware changes to bridge the auxiliary and primary memory barrier. As such, the transition to flash will not occur as quickly as in the embedded market.

A different reprogrammable application where flash memory can challenge the best current approach is data-logging systems. Whether the system is located remotely in a power substation or held by an inventory clerk in-

hand and, these systems collect data for a specific purpose. Jeopardizing that purpose by relying on magnetic media in a hostile environment or by depending on batteries for compactness makes little sense. Flash offers true nonvolatility.

CONCLUSIONS

Flash memory will be the main direct-access memory of the future. It crosses both the primary / auxiliary storage and volatile / nonvolatile boundaries. By combining reliability, density and nonvolatility in a random access memory component, Intel's ETOX flash technology approach overcomes the deficiencies of DRAM backed by hard disk, SRAM backed by battery, and EEPROM. As designers move to capitalize on flash advantages, we will see an evolution in computer architecture.

1. Myers, Glenford J., Advances in Computer Architectures, John Wiley and Sons, New York, 1982

2. Triebel, Walter A., Handbook of Semiconductor and Bubble Devices, Prentice Hall, Inc., Englewood, 1982

3. Dennings, P. J. "Third Generation Computing Systems," Computing Surveys, Vol 3, No. 4, December 1971, pp. 175-216 (c) ACM, Inc.

4. Freiberger, Paul and Michael Swaine, Fire in the Valley: The Making of the Personal Computer, Osborne-McGraw, 1984